12

ORC 77-26
OCTOBER 1977

ADAPTIVE FORECASTING WITH AN AR(1) MODEL

by
ROBERT M. OLIVER

OPERATIONS RESEARCH CENTER

UNIVERSITY OF CALIFORNIA · BERKELEY

E R R A T A

ADAPTIVE FORECASTING WITH AN AR(1) MODEL

by

Robert M. Oliver
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

1. page 7: (i) ... $\alpha_t = \alpha_t\left(z_t^2, \sigma_a^2, \sigma_\phi^2\right)$

2. page 9: Third paragraph, second line: ... $\sigma_\phi^2 = 0$, $\alpha_t = 1$.

   Delete $\beta_t(z_{t+1}/z_t)^2$.

3. page 11: First paragraph. Replace 5th sentence by:

   The coefficient of the forecast residual in (14a) does not depend on past observations (as does α_t) but rather on the most recent a priori estimate of variances.

4. page 12: Equation (17): $k_t = \dfrac{\left(\sigma_\phi^2 + \sigma_c^2\right)}{\sigma_a^2 + z_t^2\left(\sigma_\phi^2 + \sigma_c^2\right)}$; $e_{t+1} = z_{t+1} - \mu_\phi z_t$.

6 ADAPTIVE FORECASTING WITH AN AR(1) MODEL

by

10 Robert M. Oliver
Department of Industrial Engineering
and Operations Research
University of California, Berkeley

11 Oct 77

12 20p.

9 Research rept.,

14 ORC-77-26

15 [illegible]

OCTOBER 1977 ORC 77-26

This research has been partially supported by the Office of Naval Research under Contract N00014-76-C-0134 with the University of California. Reproduction in whole or in part is permitted for any purpose of the United States Government.

270 750

10 12 017

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER
ORC 77-26

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle)
ADAPTIVE FORECASTING WITH AN AR(1) MODEL

5. TYPE OF REPORT & PERIOD COVERED
Research Report

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
Robert M. Oliver

8. CONTRACT OR GRANT NUMBER(s)
N00014-76-C-0134

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Operations Research Center
University of California
Berkeley, California 94720

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
NR 047 033

11. CONTROLLING OFFICE NAME AND ADDRESS
Office of Naval Research
Department of the Navy
Arlington, Virginia 22217

12. REPORT DATE
October 1977

13. NUMBER OF PAGES
18

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Forecasting
Adaptive Forecasting
Exponential Smoothing
Filtering

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

(SEE ABSTRACT)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ACKNOWLEDGEMENTS

In a recent visit to New Zealand and Australia I had the opportunity to give several talks on topics related to exponential forecasting and Kalman filter models. In a small way this paper expresses my thanks to the Office of Naval Research for supporting my travels to these countries to attend their National Operations Research Conferences, to Professor Hans Daellenbach, Mervyn Rosser and Andrew Smith at the University of Auckland and Canterbury, to Andrew Smith at the Department of Health in New Zealand, to Mike Chapman and Colin Jex who raised several questions and pointed out several technical difficulties, to David Lee at the South Australia Institute of Technology and Professor R. B. Potts at the University of Adelaide who initiated plans for the trip and has been a source of inspiration on many occasions.

ABSTRACT

Updating formulas for the forecasts of a one-parameter autoregressive model are obtained when the parameter is assumed random. It is shown that the updated forecasts are similar to those derived from exponentially weighted moving average forecasts with the important difference that forecasts can lie outside the interval containing the old forecast and the new observation. Based on the "growth" of the new observations the updated confidence intervals may become larger or smaller than the old ones. Similarities to and differences between a Box-Jenkins model, a Kalman Filter and a model proposed by Makridakis and Wheelwright are illustrated.

1. INTRODUCTION

In several industrial and government problems to which I have been exposed, practitioners suggested that some sort of exponentially weighted moving average (EWMA) forecast should be used even though it was recognized that the underlying structure of the process was not described by an ARIMA (0,1,1) model (see Box and Jenkins [1976], Brown [1961]).

As is well known the EWMA models have the appealing and practical feature that a forecast ℓ periods into the future from an origin at t can be expressed as a weighted moving average of the historical sequence of past observations. Thus the forecast at t can be expressed as a linear interpolation of the previous forecast and the new observation

$$\begin{aligned} \hat{z}_t(\ell) &= E[z_{t+\ell} \mid z_t, z_{t-1}, z_{t-2}, \ldots] = E\left[z_{t+\ell} \mid H_t^{(z)}\right] \\ &= \hat{z}_t(1) = (1-\theta)\hat{z}_{t-1}(1) + \theta z_t \qquad \ell \geq 1\ ,\ |\theta| < 1\ . \end{aligned} \tag{1a}$$

The confidence intervals in these forecasts can also be estimated from

$$\hat{v}_t(\ell) = \mathrm{Var}\left[z_{t+\ell} \mid H_t^{(z)}\right] = (1 + (\ell - 1)(1-\theta)^2)\sigma_a^2 \tag{1b}$$

where σ_a^2 is the variance of the stationary noise distribution.

Some of the features of (1) which appeal to practitioners are the minimal data storage requirements, the fact that new data is introduced into the updated forecasts by weighting the most recent observation, that distant historical observations are weighted less heavily than recent data and that the forecasts are based on a first-order moving average equation of motion.

If one uses Box and Jenkins methods for forecasting autoregressive and moving average processes (as in the EWMA model), parameters are estimated by any one of several procedures and then assumed to be fixed throughout an interval of time in which forecasts are used. One is led to ask the natural question: Should the parameters, such as θ in (1), be updated and if so, how frequently? In the model presented in this paper we assume that the parameter itself is a random variable and is reestimated as frequently as forecasts are revised. The results obtained from a simple adaptive model in which one requires that parameters be updated hand in hand with forecasts gives numerous insights into how one might deal with changes or discontinuities in the data.

In the developments that follow we have occasion to apply Bayes Theorem to a formula for updating a Gaussian distribution based on the observation of the random variable x. If we assume that

$$x \mid m \sim N\left(m, \sigma_x^2\right) \; ; \; m \sim N\left(\mu_m, \sigma_m^2\right)$$

it follows from Bayes Theorem (De Groot [1976]) that the conditional distribution of m given the observation x is

$$m \mid x \sim N\left(\mu_m', \sigma_m'^2\right)$$

with updated mean and variance given by

$$\text{(2a)} \qquad \mu_m' = \mu_m + \frac{\sigma_m^2}{\sigma_x^2 + \sigma_m^2}(x - \mu_m) = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_m^2}\mu_m + \frac{\sigma_m^2}{\sigma_x^2 + \sigma_m^2}x$$

and

$$\sigma_m'^2 = \frac{\sigma_x^2\sigma_m^2}{\sigma_x^2 + \sigma_m^2} . \tag{2b}$$

Equation (2a) has the feature that the new expectation μ_m' is a linear interpolation of the old expectation and the new data. If σ_m^2 is large the new data x, is weighted heavily, otherwise μ_m' is close to μ_m. In (2b) $\sigma_m'^2 < \sigma_m^2$ due to the fact that a new piece of data has been obtained and one's new estimate of m is more precise.

2. THE FORECASTING MODEL

To deal with a situation where the parameter is not known with certainty, we make the following assumptions regards the underlying equation of motion for the time series $\{z_t\}$,

(A1) $$z_{t+1} = \phi_{t+1} z_t + a_{t+1} \qquad t = 0,1,2, \ldots ,$$

(A2) $$a_t \sim N\left(0, \sigma_a^2\right) \qquad \text{iid}$$

(A3) $$\phi_{t+1} = \phi_{t+1} \mid H_t^{(z)} \sim N\left((\mu_\phi)_{t+1}, \left(\sigma_\phi^2\right)_{t+1}\right) .$$

(A1) is the equation of motion for an autoregressive model of order 1. a_{t+1} is a random shock at time t and distributed according to (A2). ϕ_{t+1} is the autoregressive parameter whose distribution prior to observing z_{t+1} is given by (A3), where the moments $(\mu_\phi)_{t+1}$ and $\left(\sigma_\phi^2\right)_{t+1}$ depend in a known way on the history of the z_t process: $H_t^{(z)}$. In both (A2) and (A3) the random variables are Gaussian.

The timing of events is as follows: based on a prior estimate of the mean of ϕ_{t+1} we make the one period forecasts

(3a) $$\hat{z}_t(1) = E[z_{t+1} \mid z_t] = \mu_\phi z_t$$

and

(3b) $$\hat{v}_t(1) = \text{Var}\,[z_{t+1} \mid z_t] = z_t^2 \sigma_\phi^2 + \sigma_a^2 ,$$

where, for simplicity, the subscript on μ_ϕ and σ_ϕ^2 is temporarily deleted. The observation z_{t+1} occurs and is recorded. Based on this observation we obtain a new history $H_{t+1}^{(z)}$ and then use Bayes Theorem to

reestimate the ϕ distribution with updated parameters $(\mu_\phi')_{t+1}$ and $\left(\sigma_\phi'^2\right)_{t+1}$. Note that this problem is not as trivial as it may seem for the simple reason that the realized value of z_{t+1} is affected by two random variables a_{t+1} and ϕ_{t+1}. The algorithm which computes $(\mu_\phi')_{t+1}$ and $\left(\sigma_\phi'^2\right)_{t+1}$ must decide how to allocate the forecast residual $z_{t+1} - \hat{z}_t(1)$ to random occurrences in both a_{t+1} and ϕ_{t+1}. We now make assumption

$$\text{(A4)} \qquad E\left[\phi_{t+2} \mid H_{t+1}^{(z)}\right] = E\left[\phi_{t+1} \mid H_{t+1}^{(z)}\right] = (\mu_\phi)_{t+2} = (\mu_\phi')_{t+1} = \mu_\phi' \ .$$

In other words the realization of ϕ_{t+2} conditional on the historical observations $z_{t+1}, z_t, \ldots$ has a prior expectation equal to its most recent posteriori expectation. We are then in a position to recompute the next period forecast as

$$\text{(4a)} \qquad \hat{z}_{t+1}(1) = E[\phi_{t+2} z_{t+1}] = \mu_\phi' z_{t+1} \ .$$

In the special case where $\phi_1 = \phi_2 \cdots = \phi$, a fixed but unknown quantity, ϕ_{t+2} will be distributed according to the posterior distribution of ϕ_{t+1}, i.e., $\phi_{t+2} \sim N\left((\mu_\phi')_{t+1}, \left(\sigma_\phi'^2\right)_{t+1}\right)$. Thus, we also have

$$\text{(4b)} \qquad \hat{v}_{t+1}(1) = z_{t+1}^2 \sigma_\phi'^2 + \sigma_a^2 \ .$$

One of the main results of this paper is to show how to proceed from the forecasts in (3) to the revised forecasts in (4) without need to explicitly compute the updated parameters. In fact μ_ϕ' is never explicitly required even though it is carried implicitly throughout the analyses. The

derivation of results for the revised forecast is straightforward. It follows from (A1) and (A2) that conditionally on ϕ_{t+1} the growth in z_t is

(5a) $$\frac{z_{t+1}}{z_t} \Big| \phi_{t+1} \sim N\left(\phi_{t+1}, z_t^{-2}\sigma_a^2\right) .$$

With ϕ_{t+1} substituted for m, $z_t^{-2}\sigma_a^2$ for σ_x^2, $(\mu_\phi)_{t+1}$ for μ_m, $\left(\sigma_\phi^2\right)_{t+1}$ for σ_m^2 in (1) and (2) and Assumption (A3) the posterior distribution of ϕ_{t+1} conditional on the newly observed growth is

(5b) $$\phi_{t+1} \Big| \frac{z_{t+1}}{z_t} \sim N\left((\mu_\phi')_{t+1}, \left(\sigma_\phi'^2\right)_{t+1}\right)$$

where updated μ_ϕ' and $\sigma_\phi'^2$ (subscripts deleted) are now given by

(6a) $$\begin{aligned} \mu_\phi' &= \alpha_t\mu_\phi + (1 - \alpha_t)\,\frac{z_{t+1}}{z_t} \qquad & z_t \neq 0 \\ &= \mu_\phi & z_t = 0 \end{aligned}$$

(6b) $$\sigma_\phi'^2 = \alpha_t\sigma_\phi^2 < \sigma_\phi^2$$

with

(6c) $$\alpha_t = \sigma_a^2\left(\sigma_a^2 + z_t^2\sigma_\phi^2\right)^{-1} .$$

To obtain the revised forecast of z_{t+2} at origin $t + 1$ in terms of the old at t we substitute μ_ϕ' for μ_ϕ and $t + 1$ for t in (4) and (6) to obtain

$$\hat{z}_{t+1}(1) = \mu'_\phi z_{t+1}$$

(7) $$= ((\alpha_t \hat{z}_t(1)) + (1 - \alpha_t) z_{t+1}) \left(\frac{z_{t+1}}{z_t}\right) \qquad z_t \neq 0$$

$$= \hat{z}_t(1) \qquad z_t = 0 \; .$$

Again, for the special case where $\phi_1 = \phi_2 = \cdots = \phi$ the updated variance is

(8) $$\hat{v}_{t+1}(1) = \sigma_a^2 + z_{t+1}^2 \sigma_\phi'^2 \; .$$

What insights can one obtain from such a model and the updating formula for new forecasts? It seems to me there are several worth discussing:

(i) The weights $\alpha_t = a_t\left(z_t^2, \sigma_a^2, \sigma_\phi^2\right)$ are data dependent. They also depend on the uncertainty in the noise $\left(\sigma_a^2\right)$ and the parameter estimate $\left(\sigma_\phi^2\right)$ but are independent of μ_ϕ, the expectation of the parameter.

(ii) The revised value for the expectation of the autoregressive parameter, μ'_ϕ, is a linear interpolation of the old expectacion μ_ϕ and the most recent growth measured as the ratio of two successive observations, z_{t+1}/z_t. The parameter itself is revised by a rule similar to one used in EWMA models.

(iii) The revised value for the new forecast is a linear interpolation of the old forecast and the new observation multiplied by the growth z_{t+1}/z_t. Thus, an extremely large or extremely small value of z_{t+1}/z_t can in one time period force the revised forecast to lie outside the

interval $(z_{t+1}, \hat{z}_t(1))$ or $(\hat{z}_t(1), z_{t+1})$ until the parameters and forecasts readjust in the periods that follow.

(iv) As the estimate of ϕ becomes more certain and σ_ϕ^2 becomes small, α_t approaches unity and $\hat{z}_{t+1}(1) \rightarrow \hat{z}_t(1) \frac{z_{t+1}}{z_t} = \mu_\phi z_{t+1}$.

(v) Forecasts and confidence intervals can be initiated with little or no knowledge about the historical behavior of ϕ_t . Furthermore, if one suspects from external considerations that the uncertainty in one's knowledge of μ_ϕ has increased one can immediately insert this new information into the weight α_t and the formulas for the revised forecasts.

3. COMPARISONS WITH OTHER MODELS

The simplest one-parameter autoregressive model (AR(1)) has an equation of motion (Box and Jenkins [1976])

$$(9) \qquad z_{t+1} = \phi z_t + a_{t+1} \qquad |\theta| < 1 \ .$$

In other words (9) is based on Assumptions (A1) and (A2) with ϕ assumed given and fixed. It is well known that future forecasts and confidence intervals are given by

$$(10) \qquad \hat{z}_t(\ell) = \phi^{\ell} z_t \qquad \ell \geq 1$$

$$(11) \qquad \hat{v}_t(\ell) = \sigma_a^2 \frac{1 - \phi^{2\ell}}{1 - \phi^2} \ .$$

As we have already pointed out, when $\phi_t = \phi$ is known with certainty in (A1) - (A3) we have $\sigma_\phi^2 = 0$, $\alpha_t = 1$, $\beta_t(z_{t+1}/z_t)^2$. Equations (7) and (8) reduce to

$$\hat{z}_t(1) = \hat{z}_{t-1}(1) \frac{z_t}{z_{t-1}} = \phi z_t$$

$$\hat{v}_t(1) = \sigma_a^2$$

in agreement with (10) and (11).

A simple version of a Kalman filter (Bryson and Ho [1969]) with (9) as its underlying equation of motion (system equation) assumes that, in addition to (10), there is a noisy measurement or obervation equation of the form

$$y_{t+1} = z_{t+1} + b_{t+1}$$
(12)
$$b_t \sim N\left(0,\sigma_b^2\right) .$$

z_{t+1} is never directly observed but rather is estimated through the direct observation of y_{t+1} in (12). One can therefore make a prior (before measurement of y_{t+1}) forecast of z_{t+1} and a postèrior (after measurement) estimate of the expected or likely value of z_{t+1} when it occurs. If we define the a priori and a posteriori forecasts by $\bar{z}_t(1)$ and $\hat{z}_t(1)$ and variances by $\bar{v}_t(1)$ and $\hat{v}_t(1)$ respectively, we obtain

(13a) $$\bar{z}_t(1) = \phi\hat{z}_{t-1}(1)$$

(13b) $$\bar{v}_t(1) = \phi^2\hat{v}_{t-1}(1) + \sigma_a^2 .$$

Following the measurement of y_{t+1} in (12), we again invoke Bayes Theorem to obtain the result

(14a) $$\hat{z}_t(1) = \bar{z}_t(1) + \frac{\hat{v}_t(1)}{\sigma_b^2}\,(y_{t+1} - \bar{z}_t(1))$$

(14b) $$\hat{v}_t(1) = \left((\bar{v}_t(1))^{-1} + \frac{1}{\sigma_b^2}\right)^{-1} .$$

In these expressions, $\frac{\hat{v}_t(1)}{\sigma_b^2}$ is conventionally referred to as the "Kalman Gain" while the residual $y_{t+1} - \bar{z}_t(1)$ is the "innovation." As is well known this Kalman filter provides an a posteriori forecast of "where you were" which is a linear interpolation of the a priori

forecast ("where you thought you would be") and the observation y_{t+1} . The measurement equation reduces the uncertainty but the equation of motion tends to increase it. It is clear that there are certain similarities between the updating formulas of (13) and (14) and those of (7) and (8). However, the differences are important enough to mention: In this Kalman filter model the parameter σ_ϕ^2 does not appear since ϕ is assumed known. The weight γ_t in (14) does not depend on past observations (as do α_t and β_t) but rather on the most recent a priori estimate of variances. Furthermore the rule for revising forecasts depends on σ_b^2 which is assumed known. Neither (10) nor (13), (14) explicitly allow for growth in z_t to affect the estimate of ϕ .

During the period in which an early draft of this paper has been read by several associates it has come to my attention that the AR(1) forecasting model can also be viewed as a special case of a Kalman filter suggested by Harrison and Stephens [1976] in which the underlying equation of motion is assumed to be given in terms of the parameter (rather than the state variable z_t) as

$$(15) \qquad \phi_{t+1} = \phi_t + c_{t+1} \qquad c_t \sim N\left(0,\sigma_c^2\right)$$

and the "noisy" measurement equation for the parameter is written in terms of z_t as

$$z_{t+1} = \phi_{t+1} z_t + a_{t+1} \qquad a_t \sim N\left(0,\sigma_a^2\right)$$

with z_t given by the previous measurement. Using Assumptions (A1) - (A4) and the application of Bayes Theorem we have

$$\phi_{t+1} \sim N\left(\mu_\phi, \sigma_\phi^2\right)$$

$$z_{t+1} \mid \phi_{t+1} \sim N\left(\phi_{t+1} z_t, \sigma_a^2\right)$$

$$\phi_{t+1} \mid z_{t+1} \sim N\left(\mu_\phi', \sigma_\phi'^2\right)$$

with

(16) $$\mu_\phi' = \mu_\phi + k_t e_{t+1} z_t$$

and

(17) $$k_t = \frac{\left(\sigma_\phi^2 + \sigma_c^2\right)}{\phi_a^2 + z_t^2\left(\sigma_\phi^2 + \sigma_c^2\right)} .$$

An interesting version of (6a) is the revised parameter estimate

(18)
$$\begin{aligned} \mu_\phi' &= \mu_\phi + (1 - \alpha_t)\left(\frac{z_{t+1}}{z_t} - \mu_\phi\right) \\ &= \mu_\phi + (1 - \alpha_t) e_{t+1} z_t / z_t^2 \\ &= \mu_\phi + \left(\frac{\sigma_\phi^2}{\sigma_a^2 + z_t^2 \sigma_\phi^2}\right) e_{t+1} z_t \end{aligned}$$

which also states that the correction term one adds to μ_ϕ to get μ_ϕ' is (approximately) proportional to the product of the most recent forecast residual and the previous observation of the time series.

If we interpret k_t in (17) or $z_t^{-2}(1 - \alpha_t)$ as a data-dependent factor, then (18) is, except for the structure of this factor, similar to the proposals of Makridakis and Wheelwright [1976]. They suggest that forecasts should be written in the form

$$\hat{z}_{t+1} = \phi z_t$$

$$\hat{z}_{t+2} = \phi' z_{t+1}$$

with

(19) $$\phi' = \phi + \kappa e_{t+1} z_t$$

and κ a "training constant" or "learning factor" derived from experience with the data. What is appealing about formulas (17), (18) and (19) is that corrections are proportional to the product of forecast error and the value of the "state" variable, a concept which has been applied successfully in many important mechanical and electrical guidance control problems. What is present in (17), (18) but missing from the Makridakis-Wheelwright formulation is that the "training constant" should be data and noise dependent, i.e., it may be large or small depending on values of σ_ϕ^2, σ_a^2 and z_t^2. Further results for p^{th} order autoregressive processes are described in Nau and Oliver [1978].

BIBLIOGRAPHY

Box, G. E. P. and G. H. Jenkins [1976], TIME SERIES ANALYSIS, Holden-Day.

Brown, R. G. and R. F. Meyer [1961], "The Fundamental Theorem of Exponential Smoothing," Operations Research, Vol. 9, p. 673.

DeGroot, M. H. [1976], OPTIMAL STATISTICAL DECISIONS, McGraw Hill.

Harrison, P. J. and C. F. Stevens [1976], "Bayesian Forecasting," JRSS, Vol. 38, No. 3, pp. 205-247.

Makridakis, S. and S. C. Wheelwright [1976], "Adaptive Filtering: An Integrated ARIMA Filter for Time Series Forecasting," Operations Research, Vol. 28, No. 2, pp. 425-437.

Nau, Robert F. and Robert M. Oliver [1978], "Makridakis-Wheelwright Adaptive Filtering Revisited," ORC 78-11, Operations Research Center, University of California, Berkeley, July.